The AIR Professional File

Spring 2022 Volume

Supporting quality data and decisions for higher education.

LETTER FROM THE EDITORS

In his remarks at the closing plenary session of the 2012 AIR Forum, Pat Terenzini called attention to "the dramatic transformations in information technologies and analytical power" affecting institutional research (IR) professionals. These transformations continue today, 10 years later, at a rapid pace, thanks in part to advancements in the adjacent field of data science. While IR professionals need to understand the landscape of information technologies and analytic tools that exist, perhaps more important is the ability to apply appropriate and effective methods for generating meaning and answers from the vast amounts of data available. The two articles in this issue describe social science research methods for helping to do just that.

Paul Sturgis, Leslie Marchand, David Miller, Wei Xu, and Analia Castiglioni walk us through generalizability theory (G-theory) and provide a case study that illustrates its application to analysis of learning outcomes assessment data. They propose that G-theory is useful for empirically determining the extent to which variance in an outcome measuring student performance, such as an exam score, is due to extraneous factors, such as differing grading approaches or exam versions. As the authors explain, the method allows for multiple sources of error to be separately identified and estimated in order to increase the dependability of the outcome measurement.

Haroon Atcha provides an overview of process tracing which can help answer those challenging but important "why" and "how" questions that IR professionals are often asked. While generalizability theory can be helpful for the specific yet common task of measuring student learning outcomes, process tracing can be useful for examining the effectiveness of broad university or school-level initiatives. In process tracing, the researcher examines complex chains of events by synthesizing various forms of evidence—including quantitative and qualitative—and incorporating the researcher's prior assumptions. The result is a "thicker" measurement of the concept and context-specific insights, both valuable features of institutional research.

ISSN 2155-7535

Table of Contents

Generalizability Theory and Its Application to Institutional Research

Paul W. Sturgis, Leslie Marchand, M. David Miller, Wei Xu, and Analia Castiglioni

About the Authors

Paul W. Sturgis and Leslie Marchand are statistical analysts at the University of Central Florida's College of Medicine. M. David Miller is professor of research and evaluation methodology, as well as director of collaborative assessment and program evaluation services in the School of Human Development and Organizational Studies in Education at the University of Florida. Wei Xu is psychometrician at the American Board of Anesthesiology. Analia Castiglioni is professor in the Department of Medical Education, and medical director of the Clinical Skills and Simulation Center, as well as assistant dean for clinical skills and simulation at the University of Central Florida College of Medicine.

Abstract

This article introduces generalizability theory (G-theory) to institutional research and assessment practitioners, and explains how it can be utilized to evaluate the reliability of assessment procedures in order to improve student learning outcomes. The fundamental concepts associated with G-theory are briefly discussed, followed by a discussion of the software needed to conduct a generalizability study (G-study) analysis. The article then presents a case study of a G-study analysis; this case study was conducted in order to evaluate the generalizability and dependability of an exam that third-year medical school students complete. The conclusion discusses several situations that institutional research and assessment practitioners are likely to encounter where G-theory can be used to evaluate and improve their assessment procedures in pursuit of improving student learning outcomes.

Keywords: generalizability theory, decision study, reliability, assessment

The AIR Professional File, Spring 2022

Article 156

https://doi.org/10.34315/apf1562022

INTRODUCTION

Few institutional research and assessment professionals would argue that analyzing data in support of improving student learning outcomes is not central to their mission. In fact, this "student-focused paradigm for decision support" is explicitly recognized by the Association for Institutional Research's (AIR) statement of aspirational practice (Swing & Ross, 2016, p. 3). One of the ways that institutional research and assessment practitioners can improve student learning outcomes is by learning new skills and data analysis techniques, particularly skills that will allow them to more effectively analyze data, and to explain the results of that analysis to decision-makers.

The purpose of this article is to introduce generalizability theory (G-theory) to a new audience, and to explain how it can be used to improve assessment procedures in pursuit of improving student learning outcomes. This article will first briefly discuss the fundamental concepts associated with G-theory, and then discuss the software necessary to conduct a G-study. The article will then present the results of a G-study that was conducted in order to evaluate the generalizability and dependability of an exam that third-year medical school students complete. Finally, the article concludes with a discussion of how institutional research and assessment practitioners can utilize G-theory to evaluate and improve their assessment procedures in pursuit of improving student learning outcomes.

INTRODUCTION TO G-THEORY

G-theory is an extension of, and builds on, classical test theory (CTT). In CTT, the observed measurement is composed of true measurement and random error (Brennan, 2011; Sawaki, 2012; Teker et al., 2015; Willse, 2012). Stated more formally, in CTT "X = T + E, where X represents an observed score, T represents true score, and E represents error of measurement" (Willse, 2012, p. 150). As an example, a student's score on an exam (X) is equal to their true score (T) plus any errors associated with the exam (E). The error term (E) includes all sources of error, including such things as the day of the exam, the time of the exam, the consistency with which the rater(s) are evaluating the exam, and so on. The primary advantage of G-theory as compared to CTT is that multiple sources of error can be explicitly identified and estimated (Bloch & Norman, 2012; Sawaki, 2012; Teker et al., 2015). To return to our example above, this means that the unique amount of variance that various factors associated with the exam (e.g., the individual case and the number of raters evaluating the exam) can be estimated in a generalizability study (G-study), which of course cannot be done using the CTT framework. When comparing the two approaches, Mushquash and O'Connor (2006, p. 542) stated, "G theory is a more encompassing, informative, and useful alternative."

G-theory also builds on the familiar statistical concept of analysis of variance (ANOVA) (Sawaki, 2012; Teker et al., 2015). In fact, the variance components in a G-study are typically estimated by fitting a random-effects ANOVA model to the data (Sawaki, 2012, pp. 534–535).

Broadly speaking, conducting a G-theory analysis is a two-step process in which first a G-study is conducted, and then a dependability study (D-study) is conducted. As Croker et al. (1988) note, the purpose of the G-study is to "identify important sources of variation in a given set of observations collected under various measurement conditions" (p. 288). In simpler terms, this means that the primary purpose of the G-study is to estimate the variance components associated with the different facets of the study, which would normally be treated as an undifferentiated error term if one were to use the CTT framework. Croker et al. go on to note that the purpose of the D-study is to "obtain information that could then guide the researcher in deciding which measurement conditions should be controlled and how many levels of each condition should be included to obtain adequate generalizability" (p. 288). This means that, if we return to the example discussed above, the purpose of the D-study is to examine such things as how adding a rater that is grading some of the exams, or adding one or more cases, impacts the generalizability of the assessment.

A researcher who is considering conducting a G-study should be familiar with terms such as "facet," "universe score," and "dependability." A facet is defined as "a systematic source of variability that may affect the accuracy of the generalization one makes" (Sawaki, 2012, p. 535). To return to our above example, one of the facets that may be of interest could be the number of raters that we are using to evaluate the assessment. Other examples of facets include the individual exam items (in our example, the case); an exam given on different days/times could be a facet as well. Similarly, a universe score is defined as "the average score a candidate would have obtained across an infinite number of testing [*sic*] under measurement conditions that the investigator is willing to accept as exchangeable with one another" (pp. 534–535). This is, of course, very similar to the "true score" in CTT. Finally, dependability is defined as "the extent to which the generalization one makes about a given candidate's universe score based on an observed test score is accurate" (p. 534). As discussed above, the ultimate goal of a G-study is to determine the dependability of a measurement. In other words, the goal is to answer a research question such as this one: "If student A received a score of 90 percent on an exam, to what extent can we be confident that their 90 percent is an accurate reflection of their knowledge and abilities?"

Another strength of G-theory is that it incorporates the concept of relative and absolute decisions, which are related to the concept of norm-referenced and criterion-referenced testing. In norm-referenced testing, which is associated with the concept of relative decisions, the focus is on "the extent to which candidates are rank-ordered consistently across test tasks, test forms, occasions, and so on" (Sawaki, 2012, p. 534). Similarly, in criterion-referenced testing, which is associated with the concept of absolute decisions, the focus is on "the extent to which candidates are consistently classified into different categories (score or ability levels) across test forms, occasions, test tasks, and so on" (p. 534). The reliability index for relative decisions is typically referred to as the generalizability coefficient (Eρ2). Likewise, the index of dependability (ϕ), which is often called the phi coefficient, is used to make absolute decisions (pp. 534–535).

SOFTWARE FOR CONDUCTING A G-STUDY

Despite the fact that G-theory has been discussed in the literature since the 1970s (Cronbach et al., 1972), for many years it was used infrequently because one needed specialty software in order to conduct a G-study. Readers that are interested in the history of software programs for conducting G-studies, or who are interested in conducting a G-study in a software program other than Statistical Package for the Social Sciences (SPSS) or Statistical Analysis Software (SAS) are encouraged to consult Bloch and Norman (2012), Huebner and Lucht (2019), Mushquash and O'Connor (2006), or Teker et al. (2015) for further information.

Regardless of the software package that will be used to conduct the analysis, the first step in conducting a G-study would be to ensure that your data file is in univariate format. If your data file is in multivariate format, then the VARTOCASES command in SPSS or the PROC TRANSPOSE procedure in SAS can be used to restructure your data file (IBM, 2011; SAS Institute, 2009). Table 1 illustrates the difference between univariate and multivariate data file formats.

Depending on the complexity of the design of the study, a G-study can be conducted in SPSS using the VARCOMP procedure, but the authors would recommend using SAS as discussed in the following section. For example, when using the VARCOMP procedure in SPSS, the highest order interaction term is confounded with residual error (Putka & McCloy, 2008), therefore the VARCOMP procedure obviously cannot be used to estimate the variance component associated with the highest order interaction term. The authors' experience is that, when using SPSS version 25, adding the highest order interaction term to the model using the VARCOMP procedure results in an error and all variance components receive an estimate of "0."

Readers that are interested in conducting a G-study in SPSS using the VARCOMP procedure are encouraged to consult the excellent discussion by Putka and McCloy (2008) for further details. An additional reference would be the SPSS syntax handbook available from within SPSS by selecting the "Help" menu, then selecting "Command Syntax Reference."

Table 1. Multivariate vs. Univariate Format

Multivariate Format			Univariate Format		
Student_ID	**Rater 1 Score**	**Rater 2 Score**	**Student_ID**	**Rater_ID**	**Score**
1	**90**	**95**	**1**	**1**	**90**
2	**80**	**85**	**1**	**2**	**95**
3	**70**	**75**	**2**	**1**	**80**
			2	**2**	**85**

Note: Adapted from Putka & McCloy (2008, p. 1).

CASE STUDY: OBJECTIVE STRUCTURED CLINICAL EXAM ANALYSIS

The purpose of this study was to evaluate the generalizability and dependability of an objective structured clinical exam (OSCE) that third-year medical students at a state university complete. An OSCE involves medical students rotating through a series of timed stations where they perform certain clinical tasks. Each station represents a separate medical case; the required tasks for each case may range from taking a patient history, to performing a physical exam, interpreting diagnostic studies or lab results, counseling a patient, and so forth. OSCEs often include the use of standardized patients (SPs), who are individuals who have been trained to portray patients with the particular signs or symptoms of a medical condition in a consistent manner. In some instances, due to the length of time it takes for all medical students to rotate through all OSCE cases, multiple SPs might be trained for the same case. Student performance at each station is scored using a checklist that is specific to the content of the relevant case. The trained SPs are usually the ones who also serve as raters and who complete the checklist for each student that they interact with or observe. For the purposes of this article, the use of the term "case" implies one station of an OSCE that includes the SP and the medical condition they are portraying. The primary purpose of this project was to determine how much of the variance on the exam was attributable to the student, to the case, and/or to the rater. An additional research question involved determining the overall generalizability of the assessment.

The design of the OSCE used six stations or cases, five raters per case (34 raters in total, meaning that not all raters rated each case), and 117 students. Based on these data, a G-study was conducted using the PROC HPMIXED procedure in SAS.[1]

The following variance components were estimated in this study:

- Student (p)
- Case (c)
- Rater (r(c))
- Student * Case (p * c)
- Student * (Rater: Case), and residual (p * (r:c))

The results of the analysis are summarized in table 2. The table illustrates how the variance components and so forth change as the number of cases increase from six to eleven. As the results demonstrate, the largest variance components were those associated with student * case (p * c), and with student * case nested within rater and the residual (p * (r:c)). It is of course not surprising that a large amount of the variance is attributable to (p * (r:c)), since that includes the residual, which accounts for all unmeasured error. However, the fact that 33.1 percent of the variance is attributable to (p * c) is a promising finding. The variance associated with this component indicates that students are learning different skills across the different cases.

However, more variance is attributable to the rater than is attributable to either the student or the case. This indicates that more of the variation in performance on the OSCE is attributable to the subjective evaluation of the raters than is ideal.

1. See appendix 1 for the SAS syntax used in this analysis.

Table 2. Generalizability and Dependability Study

Third-Year Student OSCE									
Effect	**G-study Variance Component**	**% of Variance**		**Rater = 5, Case = 6**	**Rater = 5, Case = 7**	**Rater = 5, Case = 8**	**Rater = 5, Case = 9**	**Rater = 5, Case = 10**	**Rater = 5, Case = 11**
Student	5.72	11.13		5.72	5.72	5.72	5.72	5.72	5.72
Case	4.27	8.31		0.71	0.61	0.53	0.47	0.43	0.39
Rater (Case)	7.35	14.30		0.25	0.21	0.18	0.16	0.15	0.13
Student * Case	17.03	33.13		2.84	2.43	2.13	1.89	1.7	1.55
Student * Rater (Case) and residual	17.03	33.13		0.57	0.49	0.43	0.38	0.34	0.31
Total	51.40	100.00							
			Relative Error Variance	3.41	2.92	2.55	2.27	2.04	1.86
			Absolute Error Variance	4.36	3.74	3.27	2.91	2.62	2.38
			G Coefficient	0.63	0.66	0.69	0.72	0.74	0.75
			Dependability Index	0.57	0.6	0.64	0.66	0.69	0.71

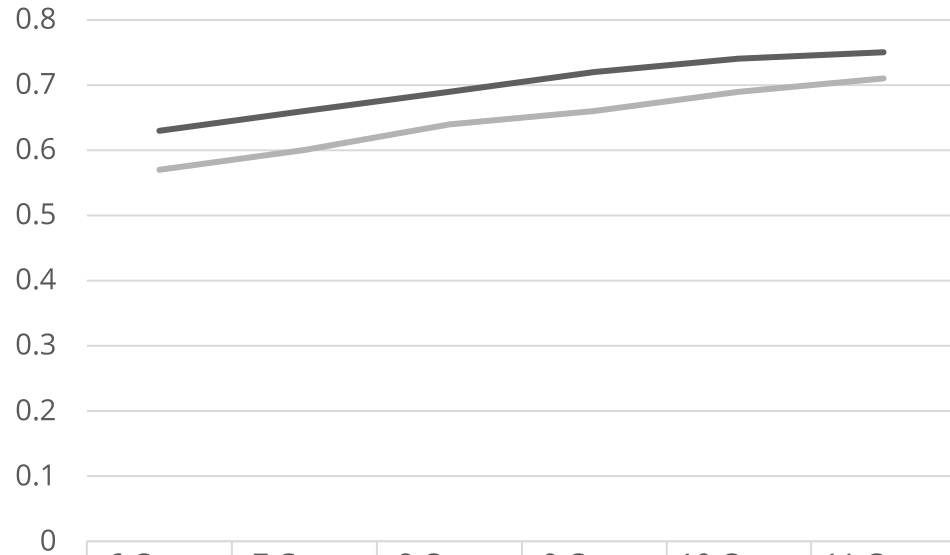
Table 2 also presents the results of the D-study.[2] As the table illustrates, as the OSCE is currently operationalized (five raters and six cases), the generalizability is .63 for relative interpretations[3] and .57 for absolute interpretations[4]. This indicates that the OSCE as operationalized is suitable for making low-stakes decisions, such as estimating student mastery of material in order to assign student grades. The generalizability of the OSCE could be increased to the .7 threshold needed for making high-stakes decisions, such as licensure or certification exams, for this type of assessment (Downing, 2004) by adding three to five additional cases. See figure 1 for additional details on the results of the D-study.

2. The D-study variance components were calculated by dividing the G-study variance component estimates by the number of cases in the study.

3. Relative error variance was calculated by summing all of the D-study variance components that include interactions with the student. The G coefficient was calculated by dividing the student variance component by the sum of the student variance component and the relative error variance.

4. Absolute error variance was calculated by summing all of the D-study variance components. The dependability index was calculated by dividing the student variance component by the sum of the student variance component and the absolute error variance. All of these calculations can easily be done in an Excel spreadsheet.

Figure 1. Generalizability Based on Number of Cases



	6 Cases	7 Cases	8 Cases	9 Cases	10 Cases	11 Cases
G Coefficent	0.63	0.66	0.69	0.72	0.74	0.75
Dependability Index	0.57	0.6	0.64	0.66	0.69	0.71

DISCUSSION AND ADDITIONAL APPLICATIONS

Although G-theory is a niche type of statistical analysis, it has many applications that those who work in institutional research and assessment are likely to encounter. The analysis that is discussed above was designed to determine how much of the variance in an exam was due to the student, the case, the rater, and so on. One of the primary findings was that, although more of the variance than is ideal is due to the raters, the majority of the variance was attributable to factors other than the raters (such as student * case and student * rater(case) and the residual), which suggests that the raters were evaluating the students' performance objectively and reliably. Those that work in institutional research and assessment, particularly those that are associated with health- and medicine-related programs, are often called on to answer these types of research questions, and hopefully the analysis presented above is useful to those researchers and can be used as a blueprint for conducting similar research projects. Those that are interested in additional ways that G-theory concepts can improve assessment procedures in the medical school curriculum are encouraged to consult Bloch and Norman's (2012) excellent discussion on the topic.

Those that work in institutional research and assessment are likely to encounter many research projects where a G-study is useful. For example, many large universities have substantial sections of writing-intensive courses where students are responding to more than one essay prompt, and where the grading is done by multiple teaching assistants. In this type of situation, G-theory can be used to determine how much of the variance in the students' scores on the essays is due to the teaching assistants, which would help to empirically determine if the teaching assistants are grading the essays in a reliable fashion. Additionally, the amount of variance that is due to the different essay prompts can be determined, which would assist faculty in making decisions about the relative difficulty of the essay prompts.

Another situation where G-theory could be useful to improve student learning is when multiple faculty members are evaluating student portfolios. Similar to the above discussion, G-theory could be used to determine the amount of variance that is due to the faculty members grading the portfolios, which would help to determine if the faculty members are grading the portfolios in a reliable fashion.

The above discussion of the possible uses for a G-theory analysis in institutional research and assessment is certainly not exhaustive. G-theory is undoubtedly a useful analytic procedure, and it can help answer many research questions related to student learning outcomes that institutional research and assessment practitioners are called on to examine.

REFERENCES

Bloch, R., & Norman, G. (2012). Generalizability theory for the perplexed: A practical introduction and guide: AMEE Guide No. 68. *Medical Teacher, 34*(11), 960–992. http://doi.org/10.3109/0142159X.2012.703791

Brennan, R. L. (2011). Generalizability theory and classical test theory. *Applied Measurement in Education, 24*(1), 1–21. https://www.tandfonline.com/doi/abs/10.1080/08957347.2011.532417

Cronbach, L. J., Gleser, G. C., Nanda, H., & Rajaratnam, N. (1972). *The dependability of behavioral measurements: Theory of generalizability for scores and profiles*. Wiley.

Croker, L., Llabre, M., & Miller, M. David. (1988). The generalizability of content validity ratings. *Journal of Educational Measurement, 25*(4), 287–299. https://onlinelibrary.wiley.com/doi/10.1111/j.1745-3984.1988.tb00309.x

Downing, S. M. (2004). Reliability: On the reproducibility of assessment data. *Medical Education, 38*:1006–1012. https://pubmed.ncbi.nlm.nih.gov/15327684/

Huebner, A., & Lucht, M. (2019). Generalizability theory in R. *Practical Assessment, Research, and Evaluation, 24*(5). https://doi.org/10.7275/5065-gc10

IBM. (2011). IBM SPSS statistics 20 documentation (pdf). https://www.ibm.com/support/pages/ibm-spss-statistics-20-documentation?msclkid=13863f66b4f211ecbf65cc95f02d060c

Mushquash, C., & O'Connor, B. P. (2006). SPSS and SAS programs for generalizability theory analyses. *Behavior Research Methods, 38*(3), 542–547. https://link.springer.com/article/10.3758/BF03192810

Putka, D. J., & McCloy, R. A. (2008). Estimating variance components in SPSS and SAS: An annotated reference guide. Human Resources Research Organization.

Salkind, N. J., Ed. (2010). *Encyclopedia of research design*. SAGE.

SAS Institute. (2009). *SAS/STAT ® 9.2: User's guide*, 2nd edition. https://support.sas.com/documentation/cdl/en/statug/63033/PDF/default/statug.pdf

Sawaki, Y. (2012). Generalizability theory. In Salkind, *Encyclopedia of research design*, pp. 534–538. https://dx.doi.org/10.4135/9781412961288

Swing, R. L., & Ross, L. E. (2016). *Statement of aspirational practice for institutional research*. Association for Institutional Research. http://www.airweb.org/aspirationalstatement

Teker, G. T., Guler, N., & Uyanik, G. (2015). Comparing the effectiveness of SPSS and EduG using different designs for generalizability theory. *Educational Sciences: Theory & Practice*, 15(3), 635–645. http://doi.org/10.12738/estp.2015.3.2278

Willse, J. T. (2012). Classical test theory. In Salkind, *Encyclopedia of research design*, pp. 534–538. https://dx.doi.org/10.4135/9781412961288

APPENDIX 1. SAS SYNTAX

```
Data rating;

infile “C:\Users\paul\Desktop\G_study.csv”

delimiter=”,” dsd;

Input ID $ Case $ Rater $ Score;

run;

ods rtf file= “C:\Users\paul\Desktop\G_study.rft”;

PROC HPMIXED method=REML;

CLASS ID Case Rater;

MODEL Score = ;

Random ID Case Rater(Case) ID*Case;

run;

ods rtf close;
```

Process Tracing for the Institutional Researcher

Haroon Atcha

About the Author

Haroon Atcha is a visiting scholar in the political science department at Arizona State University.

Abstract

Institutional researchers are often tasked with assessing why college-wide initiatives succeed or fail. This can be a difficult task: researchers need to discriminate between multiple feasible explanations, work with limited data, and produce compelling narratives. Process tracing is a qualitative methodology that enables researchers to make valid inferences in such circumstances. Process tracing focuses the researcher's attention on the sequence connecting cause and effect. It involves articulating a working theory, generating hypotheses, collecting data, assessing competing hypotheses, revising theory, and producing a narrative connecting cause and effect. This paper describes how to use process tracing for institutional research. It begins by summarizing key concepts, uses a simulated case study to give a brief overview of process tracing, discusses the importance of evidence and transparency in implementing the method, and concludes with a summary of the benefits of process tracing.

Keywords: qualitative methods, process tracing, causal inference, institutional research, institutional effectiveness

The AIR Professional File, Spring 2022

Article 157

https://doi.org/10.34315/apf1572022

INTRODUCTION

Institutional researchers are often tasked with assessing why college-wide initiatives succeed or fail (Inkelas, 2017). It can be difficult to answer "why" questions in a methodologically sound manner. Standard quantitative methods, for instance, estimate the magnitude and direction of causal effects but rarely illuminate *why* or *how* treatments cause a given outcome. Moreover, it is difficult to implement those methods when working with few and incomparable observations. Process tracing—a qualitative methodology that emphasizes the sequential links between cause and effect—can be a helpful tool in addressing these issues and assessing the success of college-wide initiatives.

I begin by summarizing key concepts: process tracing, causal process observations, diagnostic quality, and evaluating hypotheses. Next, I use a simulated case study to show how a researcher would implement process tracing in a higher education context.[1] This case study strategy includes theory generation, hypothesis generation, gathering causal process observations, assessing evidence, alternative hypotheses, and revision and completion. Following my discussion of the case study, I discuss issues of evidence and transparency. Finally, I conclude with a summary of the benefits of process tracing and stress the importance of developing a robust qualitative toolkit.

SUMMARY OF KEY CONCEPTS

Process tracing is an "analytical tool for drawing descriptive and causal inferences from diagnostic pieces of evidence—often understood as part of a temporal sequence of events or phenomena" (Collier, 2011: 824). It involves careful examination of the sequence of events that connect putative cause and subsequent effect. Process tracing involves organizing case knowledge into a cogent narrative. To organize knowledge in that way, researchers generate theory, gather evidence, test hypotheses, and reformulate theory in an iterative manner. Unlike quantitative methods, which take dataset observations as their primary form of evidence, the evidence used in process tracing is instead conceptualized as causal process observations (CPOs).

Causal process observations (CPOs) are "an insight or piece of data that provides information about context, process or mechanism, and *that contributes distinctive leverage in causal inference"* (Collier & Brady, 2004: 252). They are diagnostic pieces of evidence that allow the researcher to assess the validity of a hypothesis. Unlike dataset observations, CPOs are usually incomparable. They provide inferential value by measuring *different* variables across observations. By contrast, dataset observations provide inferential value by measuring the *same* variable across observations. CPOs provide unique insights in the assessment of causal hypotheses.

1. Given that process tracing is often used to study institutional failure (e.g., Why did our marketing campaign fail to attract more applicants?), I elect to present a simulated study rather than dive into the unflattering and personally identifying details that would be associated with a completed study. Nevertheless, I take pains to construct the case study such that key elements—data quality, availability, assessment, context, and motivating question—are representative of the experiences I have had in applying the method.

For example, we might observe that a car has been sitting outside for several months, that its paint is reddish brown, and that the paint is flaking off. These are single observations of multiple variables that are directly incomparable. Taken together, though, they provide good evidence that the car has rusted. In institutional research, CPOs may be drawn from varying units of aggregation, take different forms (e.g., documents, interview transcripts, summary statistics), or speak to different parts of a hypothesis.

Diagnostic quality refers to the distinct information a CPO brings to bear on a working hypothesis. One way of assessing the diagnostic quality of a CPO is through the framework of sufficiency and necessity. In this approach, a CPO is diagnostic to the extent that it is necessary or sufficient for the confirmation of the working hypothesis and/or disconfirmation of alternative hypotheses (Mahoney, 2012).[2] In this paper, I use a newer Bayesian framework for evaluating the diagnostic value of CPOs (Bennett, 2008; Fairfield & Charman, 2017).

Evaluating hypotheses in a Bayesian framework entails updating prior beliefs about the probability of a hypothesis being true given our CPOs. As noted by Fairfield and Charman, "We gain confidence in a given hypothesis to the extent that it makes the evidence we observe more plausible in comparison to rivals" (Fairfield & Charman, 2017: 159). To evaluate hypotheses, the researcher first articulates a prior belief about the probability of a hypothesis being true. Then the researcher can update that belief in proportion to the diagnostic value of given CPOs. In Bayesian approaches to process tracing, CPOs condition the researcher's belief in the likelihood of the working hypothesis vis-à-vis alternative hypotheses.

Process tracing is well suited for within-case analysis. This quality situates process tracing firmly within the case study tradition in higher education and institutional research.[3] Whereas regression, experimental, and quasi-experimental methods generally attempt to estimate the direction and magnitude of a causal quantity, process tracing is primarily concerned with *how* and *why* a causal effect came to be in a particular context. To assess how and why the causal effect came to be, the method uses evidence particular to that case to draw inferences about cause or lack thereof.

Process tracing is particularly useful when researchers need to diagnose initiative failure: *Why* didn't our marketing campaign increase the number of applicants? *At what points* did communication between stakeholders fail? *How* did the application process keep students from applying? Questions like these assume knowledge of a causal effect; we know the initiative failed, now we would like to know why. Process tracing provides a methodologically rigorous way of answering such questions in a robust and transparent manner.

2. In the traditional approach, CPOs could be doubly decisive if they at once confirm the working hypothesis and disconfirm alternative hypotheses, they could be a smoking gun if they confirm the working hypothesis but do not disconfirm alternative hypotheses, or they could pass a hoop test or be a straw-in-the-wind if they are necessary though not decisive and of minimal diagnostic value, respectively.

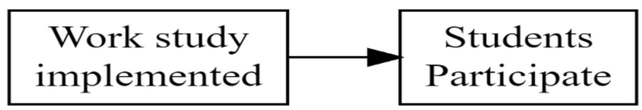
3. See Yin (2013) and Silverman (2013) for canonical and contemporary texts concerning case studies and qualitative research in general as well as Merriam (2007) for qualitative research in higher education and George Mwangi and Bettencourt (2017) for an overview of qualitative methods in institutional research.

PROCESS TRACING: THE SIMULATED CASE STUDY

Theory Generation

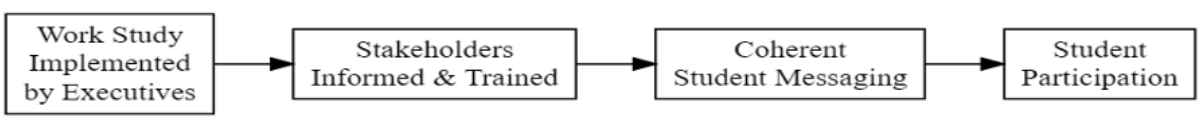
Process tracing begins by articulating both a working theory and the intermediate steps that connect cause and effect. The first step in process tracing is to clearly articulate the dependent and independent variables. As a working example, consider that we have been asked to study the implementation of a work-study initiative and to assess why student participation in the initiative is low (figure 1). Our dependent variable is student participation and our independent variable is the implementation of the initiative.

Figure 1. Work-Study Initiative Implemented



Our first task is to theorize on the steps between our independent variable and our dependent variable. Based on case knowledge and the extant literature, we might theorize that clear communication between stakeholders and students about the initiative would result in greater student participation. Consistent messaging could drive greater student awareness and participation. This expanded theory is visualized in figure 2.

Figure 2. Consistent Messaging for Greater Student Awareness and Participation



In practice, our theoretical chain of events connecting the independent and dependent variables would likely be longer and more detailed. The purpose of process tracing is to produce a *complete* narrative of the events linking the two. The example given requires various leaps between intermediate steps that would ideally be more thoroughly articulated in a full case study. As a first step, though, it provides a useful outline of the theorized process and suggests avenues for exploration.

Hypothesis Generation

After we have articulated an initial theory, our next task is to generate hypotheses that probe at the connections between steps. Confirming or failing to confirm these hypotheses should provide information about the validity of our explanation vis-à-vis alternative explanations. The theory visualized in figure 2 has three connections that we can probe for hypotheses.

First, we can ask whether stakeholders were properly informed and trained by executives. Second, we can ask whether the initiative produced coherent student messaging. Finally, we can ask whether students participated, given such information. For the purpose of illustration, consider the second hypothesis (H2):

H2: The work-study initiative produced coherent student messaging.

Assessing H2 provides information about the validity of our working theory. If CPOs support H2 and we find the initiative had coherent messaging, we have good reason to believe that low participation was not caused by poor communication. By contrast, if CPOs disconfirm H2 and we find that communication was not coherent, we would want to look more closely at the connection between executives and stakeholders to understand why communication was not coherent. H2 is useful in assessing the validity of our theoretical chain because both confirming and failing to confirm point to productive avenues for theory reevaluation.

Gathering Causal Process Observations

Having articulated a hypothesis, we next need to operationalize our measurements and gather CPOs. To do so, we need to define measurable outcomes that map onto the concepts in our hypothesis. For H2 this means answering the question, "What does *coherent student messaging* look like and how can I measure it?" Ideally, we can think of several different ways of answering that question, with the different answers probing at different parts of the concept.

For instance, we might ask stakeholders if they were aware that the college had a work-study program with a binary yes/no outcome. We could organize this information in the form of a cross-tabulation. We may also ask stakeholders to explain the goals of the program and see whether their answers are similar. If we are interviewing a large number of stakeholders, we could organize this information as free-text that could then be interpreted via topic-model. Or, if working with a small number of observations, we could use thematic or qualitative content analysis to see if themes are repeated. We may ask fellow researchers to independently code answers for theme as well, in the interest of validating our coding scheme.

Appropriate timing would also be an important aspect of coherent student messaging. Effective messaging would be timed to coincide with important milestones for applications and registration. We might investigate whether the college ran banner ads for the work-study program on its website, especially on the pages that students are likely to frequent. We would want to see whether these ads were active during the registration window. If they ran long before or long after, we might not consider that messaging to be coherent or relevant. The sequence of events would matter for assessing H2.

One of the benefits of process tracing is its ability to synthesize evidence that takes a variety of forms. Since our measurements take a variety of forms—a timeline, topic model, cross-tabulation, and so on—they are not directly comparable in that they measure different aspects of the same concept. But process tracing encourages this type of diversity when collecting evidence. Together they help us formulate a thicker measurement of the concept.

Moreover, there is no single, correct way of collecting CPOs. Researchers are free to operationalize concepts in myriad ways and should abide by best practices when implementing each. For instance, when developing a survey to ask stakeholders about their knowledge of a work-study program, researchers should take care not to design leading questions, adjust for sampling bias, and use appropriate scales.

Assessing Evidence

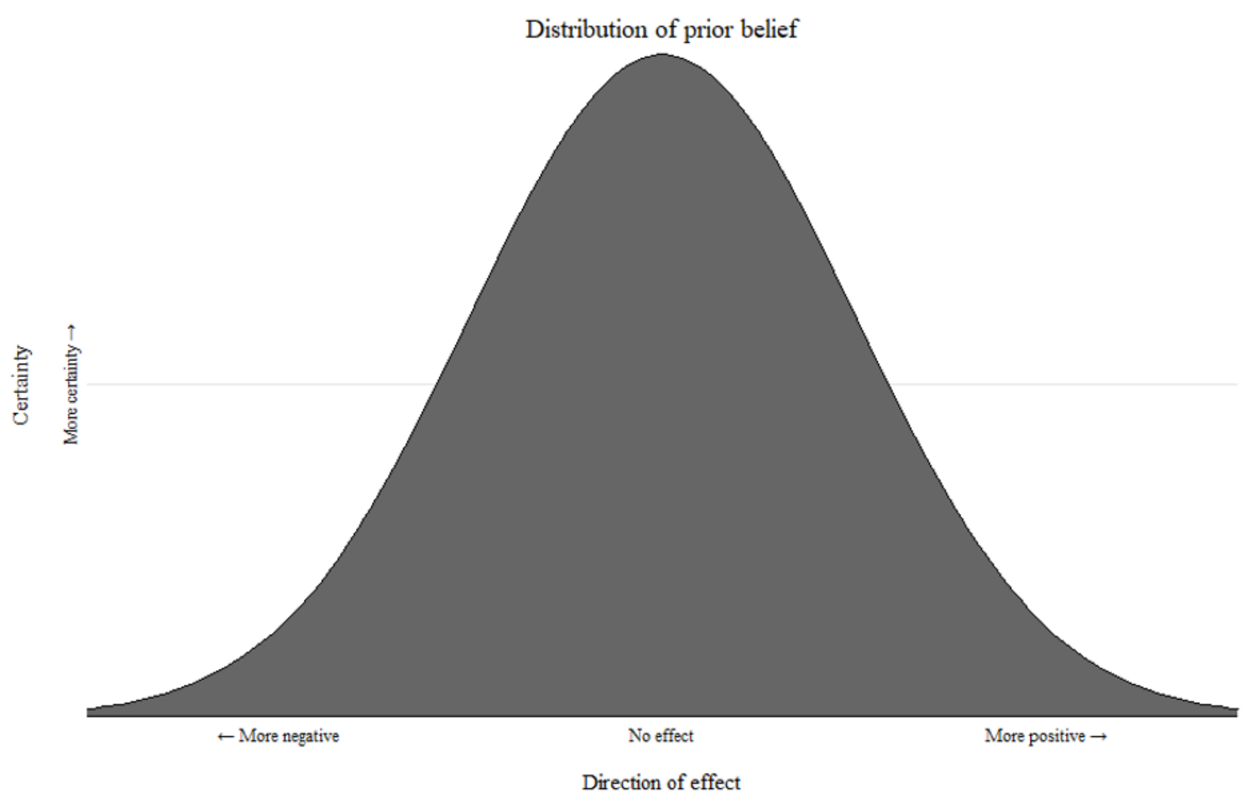
The next step is to assess the validity of our hypothesis using our collected CPOs. Consider our second hypothesis:

H2: The work-study initiative produced coherent student messaging.

To assess the validity of this hypothesis in a Bayesian framework, we first need to define our prior belief about its likelihood. Usually we would have some prior knowledge about the likelihood, but we assume for the purpose of illustration that we are starting from a point of ignorance. Our prior belief is defined by conservative expectations: extreme values are less likely than moderate ones. The most likely scenarios would involve moderately coherent or incoherent messaging; neither highly coherent nor highly incoherent messaging is likely.

This safe prior assumption can be visualized as a normal distribution across outcomes, as shown in figure 3. Knowing *nothing else*, it would be reasonable to assume that the initiative *probably* did not produce wildly coherent or wildly incoherent student messaging.

Figure 3. Normal Distribution across Outcomes



The process of describing and updating our prior belief in process tracing is not necessarily a quantitative one. The figures provided are meant as heuristic devices that model the way researchers infer from evidence. While a researcher *could* define a distribution of belief in strictly quantitative terms and *could* attach weights to CPOs that then change the density of the prior distribution, doing so would obviate many of the benefits of collecting thick, diverse measurements.

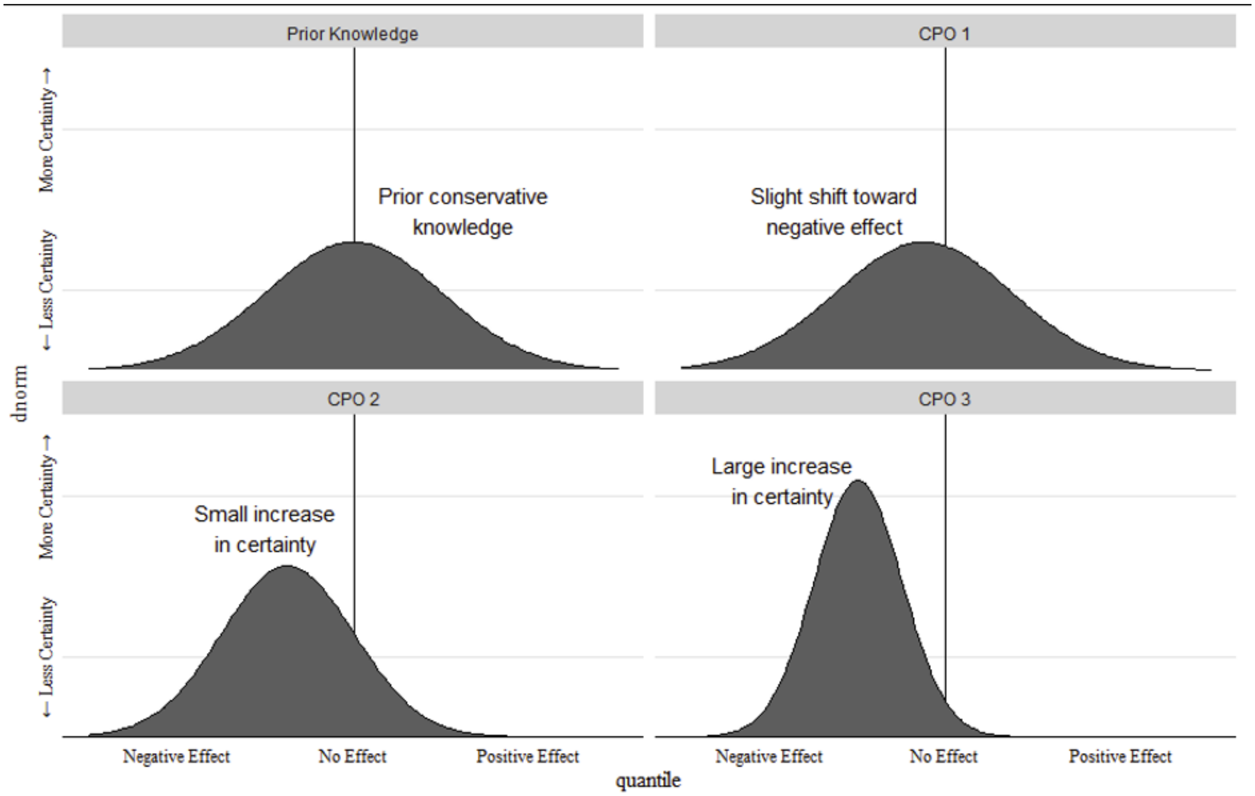
With our prior belief, we can observe CPOs and condition our belief on the evidence they provide. For illustrative purposes, we can also visualize the diagnostic effect of each CPO on our prior belief, as shown in figure 4. Consider the following CPOs:

- CPO 1: Stakeholders interviewed did not articulate a consistent understanding of what the work-study program entailed.

- CPO 2: Only a small amount of marketing material about the work-study program was specifically generated after the initiative's adoption.

- CPO 3: In interviews, multiple department heads were unaware that the college had a work-study program.

Figure 4. Diagnostic Effect of Each CPO on Our Prior Belief



Each CPO conditions our expectation about the likelihood that the work-study initiative produced coherent student messaging to varying degrees, as visualized in figure 4. What follows is an explicit articulation of how each CPO conditions our expectation. Reasoning in a clear and transparent manner is useful for several reasons. It can help increase trust in the research and can help other researchers validate our findings. In practice, this section would likely be a part of a separate transparency appendix that would be included at the end of a report rather than incorporated into the main body of the text. In the interest of demonstration, however, this material, starting in the next paragraph and concluding with the start of the next section, is included in the body of this paper.

CPO 1: Stakeholders interviewed did not articulate a consistent understanding of what the work-study program entailed. The first CPO offers minor diagnostic evidence that student communication was not coherent. If messaging were coherent, we would expect stakeholders to be able to articulate what the initiative entailed in similar terms. However, the absence of uniformity does not necessarily indicate poor communication on its own. For instance, it could be that stakeholders not engaged directly in student communications have no need to know about the initiative to do their jobs effectively. The observation of this CPO shifts our expectation slightly toward incoherent, but does little to increase our certainty.

CPO 2: Only a small amount of marketing material about the work-study program was specifically generated after the initiative's adoption. The second CPO offers more-compelling diagnostic evidence. Across all marketing material made available to students (e.g., school website, counseling webpages, course scheduling platforms, physical posters, etc.), very few mentioned the new work-study initiative after the initiative was implemented. Marketing materials did not advertise the start of the program, and did not explicitly encourage students to inquire about it; in many cases, marketing material was not updated after the initiative's implementation. These materials provide strong evidence that the initiative is not producing coherent student messaging. Again, alternative explanations may produce similar observations. For instance, it could be that the term "work-study" is intended for internal use only and that it would be marketed under a different name. Considering the first CPO, however, that alternative explanation seems unlikely. The observation of this CPO shifts our expectation more decisively toward incoherent and increases our certainty.

CPO 3: In interviews, multiple department heads were unaware that the college had a work-study program. Finally, the third CPO offers the most diagnostic evidence. Interviews with department heads revealed that many were completely unaware that the college had a work-study program. Even if a student knew to ask a college employee about the program, the college employee they ask might be unfamiliar with it and so would be unable to inform the student. This CPO provides the strongest evidence that student-communication is incoherent.Observing a fundamental misalignment across student-facing stakeholders shifts our expectation solidly to incoherent and greatly increases our certainty. Observing this CPO further recontextualizes our first CPO. The inability to consistently define work-study terminology is far more revealing of miscommunication *if we already know* that department heads were not aligned in their messaging.

In sum, these three CPOs provide strong evidence that the work-study initiative did *not* produce coherent student messaging. Even CPOs that, alone, would do little to confirm our hypothesis, help paint a more complete picture of the mechanism at work in the context of the other CPOs. For instance, observing inconsistent definitions across stakeholders is not necessarily decisive on its own; in the context of other CPOs, however, those observations help us understand the nature of the misalignment.

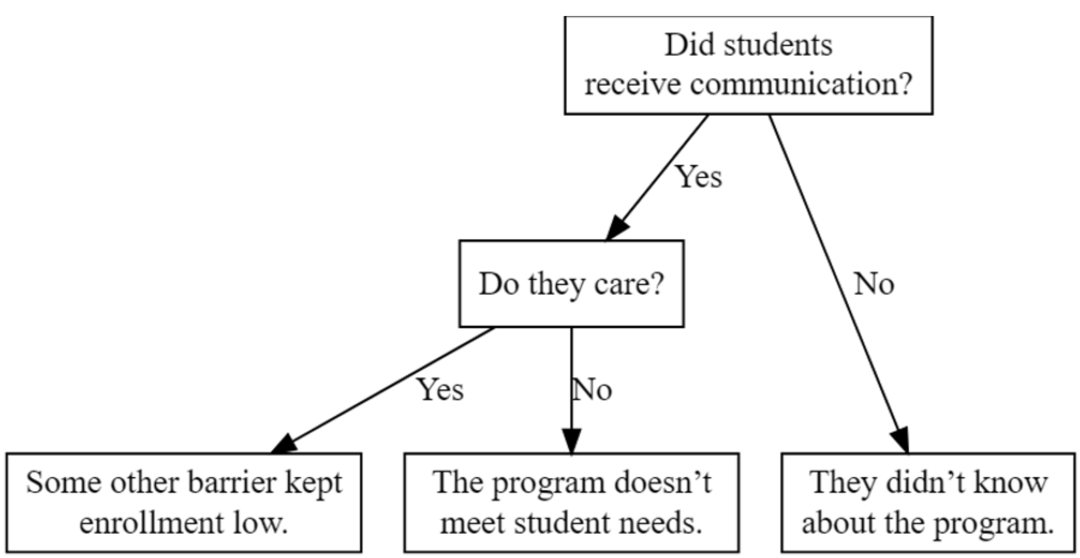
Alternative Hypotheses

Though addressing alternative hypotheses is a critical component of hypothesis testing, not all work done to dismiss alternative hypotheses is done while testing working hypotheses. Alternative explanations might not fit neatly in as direct competitors to our working hypotheses and instead might need to be addressed independently. Each CPO might provide evidence for many feasible competing explanations. Eliminating these alternative explanations is critical in producing a compelling narrative.

Generating alternative hypotheses (and the implicit counterfactual) can often be done by branching off the theoretical chain of events at different points. Branching in this manner serves at least two purposes. First, it provides a robust way of generating alternative hypotheses and explicitly acknowledging researcher assumptions. Branching off from the theoretical tree forces researchers to address the implications of mis-specified prior knowledge. Second, it is a transparent manner of articulating *which* alternative explanations the researcher chose to address and *why*.

Consider figure 5, which captures the logic of branching in the working example. Though the initial tests of our hypothesis suggest that one of the flaws in the initiative was a lack of coherent student communication, it does not necessarily follow that fixing that flaw would mend the program. For instance, we could find through deeper exploration that some students *did* know about the work-study program and that those students *still* did not enroll. Branching allows us to identify those possible alternative explanations and address them.

Figure 5. Logic of Branching in the Working Example



But research need not address *every* possible alternative explanation. Good models are parsimonious. Researchers should attempt to address all *feasible* alternative hypotheses. Determining which explanations are *feasible* is in part an iterative process of theory creation, hypothesis articulation, evidence gathering, and hypothesis assessment. If researchers choose not to address a specific alternative hypothesis, they should explain why and how the working theoretical model makes that alternative unlikely.

To return to the working example, consider our working theory, which outlines the process by which a work-study program was implemented at our hypothetical college. Though evidence suggests that poor institutional communication is at least partly responsible for low student participation, other explanations are feasible alternatives as well. For instance, it could be that work-study opportunities are offered at inconvenient times or that there are too few opportunities offered.

Addressing these alternative hypotheses entails gathering CPOs that *uniquely* speak to their validity. Again, this process of reasoning would likely be a part of a transparency appendix and not in the body of the report; it is included in-text here for illustrative purposes, however. Alternative hypotheses 1 and 2 (Ha1and Ha2) can be formalized as follows:

Ha1: Work-study opportunities are offered at inconvenient times.

Ha2: Too few work-study opportunities are offered.

As with our working hypothesis, we begin assessing the validity of these alternatives by defining measurements, collecting CPOs, and defining a prior. Instead of starting from a position of ignorance, however, assume we know that the work-study organizing committee devoted significant energy to scheduling work-study opportunities such that they would be accessible to students.

Our prior belief can reflect that institutional knowledge: we are marginally less likely to believe that work-study opportunities are offered at inconvenient times. Our normal distribution across beliefs is shifted with preference toward our prior belief. For Ha2 we begin from a position of relative ignorance since no institutional knowledge we have speaks to this dynamic.

Various CPOs condition our expectations about the likelihood of these alternatives being feasible. Consider the following CPOs that speak to the validity of our alternative hypotheses:

- CPO 4: The distribution of work-study opportunities over time (in a given day) is nearly identical to those of peer institutions.
- CPO 5: The density of work-study opportunities at a given time is proportionate to the number of students on campus at that time.
- CPO 6: Nearly all remote work-study opportunities fill to capacity.
- CPO 7: In-person work-study opportunities filled to 20 percent capacity.
- CPO 8: In-person work-study opportunities largely fell under department heads who did not know about the work-study program.
- CPO 9: Student surveys indicated that knowledge of the work-study program varied by major and department.

These CPOs complicate our theoretical explanation in that they do not neatly fit the narrative that communication *alone* is the cause of poor student enrollment. CPOs 4 and 5 provide evidence that work-study opportunities are well-aligned with student availability. Moreover, given that work-study is conceived of as an alternative to courses, it stands to reason that work-study opportunities should be provided at times that students would otherwise be taking courses. Since course density and work-study density in a given time slot are proportionate, we can reasonably infer that poor scheduling is *not* a cause of low participation.

CPOs 6 through 9 paint a more complicated picture vis-à-vis Ha2. CPOs 6 and 7 suggest that the format of the opportunity matters to students. Online work-study opportunities might be more appealing relative to in-person opportunities in the wake of COVID. However, CPO 8 suggests that the disparity across format may be a function of our main hypothesis concerning breaks in communication. CPO 9 further supports this reading. Students with majors in departments whose heads did not know about the work-study program also did not know about the program.

The cumulative weight of these CPOs supports the notion that our working hypothesis is the most feasible explanation for the observed phenomenon. Though these CPOs do not entirely dismiss the possibility that our alternative hypotheses play a role in the observed phenomenon, they place our working hypothesis higher on the continuum of feasibility than the most likely alternatives. Addressing these alternative explanations strengthens our narrative.

Revision and Completion

Ultimately, this process produces a cogent narrative of the sequence of events that connect putative cause and subsequent effect. In the process of testing hypotheses and addressing alternative explanations, we may find that our theorized linkages are insufficient for describing the phenomenon. For example, we concluded that poor communication with students is the likeliest cause of their low participation in the work-study initiative, but it is unclear why initiative creation is disconnected from stakeholder messaging. Our theory is insufficient for describing what we observe.

To address this gap, we would revise our theory, adding links, producing and testing hypotheses, and considering alternative paths that might produce similar outcomes as we did with initial theory articulation. At each step in the iterative process, we would ask the same questions as we did in the initial step: Did this step occur as hypothesized? If so, what CPOs support this interpretation? If not, what process does the evidence suggest occurred? Such questions may guide the researcher in producing a cogent, sequential account of the total process.

Ideally, the output of this process is a narrative that links putative cause and subsequent effect in an unbroken chain of events that is accessible to nontechnical audiences. The narratives produced through implementing process-tracing methodologies should be, above all, readable.

EVIDENCE AND TRANSPARENCY

Evidence and transparency warrant special attention. In particular, researchers need to articulate standards for two specific processes. First, researchers need to communicate how CPOs were collected and analyzed. Second, researchers need to communicate how they determined the diagnostic value of CPOs and why they came to the conclusions they did. The difficulty in accomplishing the first task is compounded when researchers use different types of evidence. Researchers can use several strategies to resolve this difficulty, however.

First, researchers can rely on existing standards of validity and transparency where appropriate. Often, researchers will find themselves using methods to gather and analyze evidence that have widely accepted standards of practice. Summary statistics, data visualizations, flowcharts, and regression models could be CPOs that speak to the validity of a working hypothesis. These types of evidence have well-established norms of practice, norms that researchers should abide by. For instance, researchers should, among other things, label axes, normalize and center continuous variables, provide code books, and test alternative specifications of regression models. Clearly describing how datasets were gathered and analyzed and making code available are other basic steps that are widely accepted as best practices and that increase confidence in findings.

Second, researchers ought to explicitly articulate assumptions and prior knowledge. Though prior knowledge does not often play a decisive role in making inferences when using a Bayesian framework, it is important for us to recognize that we begin research with existing knowledge. Clearly listing assumptions, especially if they are unique to the given context, also helps paint a richer picture of the processes at work. Often, transparently articulating prior assumptions can be accomplished by providing a history or context section that summarizes similar research undertaken at the college and how that research informs the researcher's assumptions.

Third, a *transparency appendix* can greatly increase the robustness of and confidence in research findings.[4] A transparency appendix is a supplemental section appended to the end of a report that includes a citation, excerpt from the cited text where appropriate, and short commentary on how this evidence supports the researcher's interpretation. In this paper I have included material that would otherwise be a part of a separate transparency appendix in-text rather than presenting it in its own section. Paragraphs exploring how a CPO was gathered and how it conditions our belief in the working hypothesis provide a template for how a transparency appendix is formulated. These need not be in the body of the text, as they are here, but they should be accessible and should explain a researcher's logic such that a reasonable reader can follow the underlying reasoning.

Transparency is a core tenet of all research but it is particularly necessary when undertaking process tracing. Since results from this method take the form of a narrative written in accessible language, leaps of logic and unsupported assumptions are more likely to catch college stakeholders' eyes. Often, results generated by regression modeling and quasi-experimental methods benefit from a degree of obscurity. Most college stakeholders are not familiar with the particularities of statistical modeling. As such, institutional researchers are unlikely to be questioned about their decision to use fixed versus

4. Active citations and transparency appendices are two tools gaining broad traction in the recent push for transparency in qualitative research. Institutional researchers producing interactive reports (with tools like Markdown, for instance) should find it relatively easy and low-cost to incorporate these tools into their research. For a brief summary of the need for and tools facilitating transparency in qualitative research for political science research, see Moravcsik (2013).

random effects or about the heteroskedastic qualities of their models' residuals. They *are* likely to face scrutiny over the decision to incorporate prior knowledge in a particular way, however. The steps listed above—using existing standards when appropriate, articulating prior knowledge and assumptions, and creating a transparency index—go a long way toward alleviating those concerns.

CONCLUSION: SUMMARY OF THE BENEFITS OF PROCESS TRACING

Process tracing is uniquely suited to the study of institutions and their development. Indeed, much of the extant literature that makes use of process tracing takes the institution as its primary unit of analysis. Seminal works of process tracing have examined the development of a nuclear taboo (Tannenwald, 1999), the development of communal riots in India (Brass, 1997), and the sequence of events leading to social revolutions (Skocpol, 2015). Though these cases might appear to be a world away from institutional research, they exemplify how process tracing can help extract analytical insights from the study of institutions.

Institutional researchers are particularly well-served by this methodology. Process tracing produces institution-specific insights, facilitates the synthesis of varied forms of evidence, and allows for the incorporation of prior knowledge.

Unlike academic research, which often needs to balance concerns over specificity and generalizability, the institutional researcher's main concern is with delivering results that speak to the specific context in which they work. To that end, process tracing—and qualitative methods more broadly[5]—lend themselves to this endeavor. A given college diverges from the average institution in myriad ways. Colleges react to fluctuations in local economic conditions, reflect the particularities of their communities, and have histories of success and failure with college-wide initiatives. Results derived through process tracing can speak directly to those particularities.

Process tracing also allows for the synthesis of varied forms of evidence. As noted by Harper and Kuh, evidence can often take the form of "observations, document analyses, and reflective journaling," as well as interviews and focus groups in learning assessment (Harper & Kuh, 2007: 11). Institutional researchers often have access to vast amounts of nonquantitative data. Process tracing facilitates the robust analysis of institutional trends with the use of survey responses, interviews, and texts. Indeed, the narratives produced by process tracing are strengthened when they are supported by varied and nuanced forms of evidence.

Finally, when approached from a Bayesian perspective, process tracing allows for the incorporation of prior knowledge. Just as it allows for production of institution-specific knowledge, so too does process tracing build on institution-specific knowledge. In this way it is similar to, but distinct from the first benefit. Various sources of prior information, like knowledge of institutional history and experience with similar programs, may point the researcher away from the extant research for good

5. A 2017 special issue of New Directions for Institutional Research focuses on just this topic, expanding on the use of qualitative methods for assessing broad-based initiatives (Inkelas, 2017), student experiences (Friedensen et al., 2017), and departmental effectiveness (Williams & Stassen, 2017). These pieces serve as helpful introductions to the use of qualitative methods but point to a relative paucity of scholarly works expanding on the actual implementation of such methods in institutional research.

reasons and toward more-robust sources of insight. A Bayesian approach to process tracing allows for the analysis of institutional change in a way that incorporates context-specific forms of evidence.

Qualitative methods are an important aspect of an institutional researcher's toolbox. Process tracing is one such tool and is especially appropriate for the analysis of large, college-wide initiatives. Process tracing allows for the analysis of complex chains of events and excels in identifying context-specific insights central to the institutional researcher's job.

WORKS CITED

Bennett, A. (2008). Process tracing: A Bayesian perspective. In J. M. Box-Steffensmeier, H. E. Brady, & D. Collier (Eds.), *The Oxford handbook of political methodology.* Oxford University Press.

Brass, P. R. (1997). *Theft of an idol.* Princeton University Press.

Collier, D. (2011). Understanding process tracing. *Political Science & Politics, 44*(4), 823–830.

Collier, D., & Brady, H. (2004). *Rethinking social inquiry: Diverse tools, shared processes.* Rowman & Littlefield.

Fairfield, T., & Charman, A. E. (2017). Explicit Bayesian analysis for process tracing: Guidelines, opportunities, and caveats. *Political Analysis, 25*(3), 363–380.

Friedensen, R. E., McCrae, B. P., & Kimball, E. (2017). Using qualitative research to document variations in student experience. *New Directions for Institutional Research, 2017*(174), 53–63.

George Mwangi, C. A., & Bettencourt, G. M. (2017). A qualitative toolkit for institutional research. *New Directions for Institutional Research, 2017*(174), 11–23.

Harper, S. R., & Kuh, G. D. (2007). Myths and misconceptions about using qualitative methods in assessment. *New Directions for Institutional Research, 136,* 5–14.

Inkelas, K. K. (2017). Using mixed-methods to assess initiatives with broad-based goals. *New Directions for Institutional Research, 2017*(174), 39–52.

Mahoney, J. (2012). The logic of process tracing tests in the social sciences. *Sociological Methods & Research,* (March 2), 570–597.

Merriam, S. B. (2007). *Qualitative research and case study applications in education: Revised and expanded from case study research in education.* Jossey-Bass.

Moravcsik, A. (2013). Transparency: The revolution in qualitative research. *Political Science & Politics, 47*(1), 48–53.

Silverman, D. (2013). *Doing qualitative research: A practical handbook.* SAGE.

Skocpol, T. (2015). *States and social revolutions: A comparative analysis of France, Russia, and China.* Cambridge University Press.

Tannenwald, N. (1999). The nuclear taboo: The United States and the normative basis of nuclear non-use. *International Organization, 53*(3), 433–468.

Williams, E. A., & Stassen, M. L. (2017). Context matters: Using qualitative inquiry to inform departmental effectiveness and student success. *New Directions for Institutional Research, 2017*(174), 65–80.

Yin, R. K. (2013). *Case study research: Design and methods.* SAGE.

THANK YOU

Authors who submit materials to AIR Publications receive feedback from members of the AIR community. AIR expresses sincere appreciation for the individuals who serve as peer reviewers.

The following individuals reviewed manuscripts recently submitted to *The AIR Professional File*.

Kristin Buscher

Rhyan Conyers

Mollie Miller

Amy Pilcher

Sharron Ronco

Annette Sieg

Suzanne Simpson

Carolyn Wiley

ABOUT *THE AIR PROFESSIONAL FILE*

The AIR Professional File features journal-length articles grounded in relevant literature that synthesize current issues, present new processes or models, or share practical applications related to institutional research. All submissions are peer-reviewed. For more information about AIR Publications, including *The AIR Professional File* and instructions for authors, visit www.airweb.org/publications.

Association for Institutional Research
Christine M. Keller, Executive Director and CEO
www.airweb.org
ISSN 2155-7535